**Data Management Plan**

**Presented to Professor Green**

**Executive Summary**

Outlined in this report is a data management plan, which includes recommendations for database management, data storage and data sharing. Best practices call for data to be made available as *open data*, and follow the FAIR (findable, accessible, interoperable and reusable) guiding principles for data management. This not only meets the university’s policy on research data, it is often required by peer reviewed journals and is shows accountability for public funds.

We have prepared this report - based on requirements outlined by the university librarian and input from the research project manager, Professor Green - to be thorough and of benefit to the research team.

**Project Description and Scope**

Dr. Green is focusing his research on evaluating how teams perform in high-pressure situations. He has selected hospital workers in primary care units as a case study for his analysis. Essentially the professor wants to know whether the high-pressure environment in which these individuals operate in affects how teams function as a cohesive unit.

The professor’s primary methodology includes interviews and written analysis. While he has only two streams of data collection, the manner in which his information is stored varies greatly. The file scope for Dr. Green is rather diverse. This is proven by the use of many file types such as Word, PDF, Excel and MP3. Essentially file types will either contain text or audio. In terms of volume the real burden will come from the text files, such as Word, PDF and plain text, as these documents number in the hundreds. The audio files, which house the interviews, are relatively few in comparison.

Dr. Green has sole ownership of the data collected through his research. Although graduate students sometimes help him, the professor is wary of disclosing his data to others. For this project he is employing two graduate students who will need to have access to some parts of the data.

The primary objective of the data management plan should focus on Dr. Green’s requirements for organization and security. Organization is needed for the classification and sorting of his text files. Additionally security is paramount to his data needs as there are confidentiality requirements for his audio files. Lastly it is important to note the professor does not want his data to be open sourced.

**Documentation, Organization and Storage**

As mentioned previously Dr. Green’s data is spread across a variety of file formats. While the text related documents out number the audio files, it can be assumed that the latter will be a higher burden in terms of file size. A MP3 file containing an hour of content, depending on bitrate quality, can range anywhere from four megabits to one hundred and forty-four megabits. In comparison an average PDF/Word file is under one megabit.

According to the up follow questions with professor, Dr. Green is currently using around twenty-four gigabytes of storage on his USB drive. He also predicts that his future needs will triple in size over the next ten years. As such the professor will be looking for a storage platform that will scale in capacity parallel with his research.

The professor is currently using three different types of storage platforms for each of the different types files he deals with. Text and excel documents are presently stored using Zotero, audio files with DropSync.com and Google Docs for transcripts. Best practices towards data management go against what the professor is currently doing. Instead Dr. Green should consider consolidating his files.

The first task should be converting any Word documents into PDF files. Not only would this reduce the number of file types but it would also prevent other people from editing the document. Secondly the professor could benefit from the use of file containers such as ZIP files. These containers can store similar-in-content files, whereby drastically reducing the number of files the professor has to sort through. An example of this would be grouping together a MP3 file, its associated transcript and any other supporting documents pertaining to the interview into a single ZIP file. Lastly ZIP files would also reduce the file sizes of the professor’s MP3 files, whereby reducing his total storage needs.

The use of ZIP files will also have a secondary benefit for Dr. Green. By utilizing this simple categorization method, the professor will be making use of metadata. Given the ZIP files will house directly linked content, each file makes use of metadata in the sense that the files will be categorized. The use of ZIP files, and by extension metadata, will ensure that the professor can easily identify and reuse the data at a later date.

Theoretically using files containers will reduce the volume of files, however Dr. Green will still need to house these files somewhere that’s secure but also accessible. It will be recommended that Dr. Green stores his file containers on a single cloud platform, rather than three that he is currently using. Additionally since his data is from Canadian sources he will have to abide by Canadian storage laws. These laws dictate that content collected in Canada must be stored in Canadian servers. Therefore based on his current and future storage needs the professor should consider utilizing cloud storage from Sync.com,

Sync.com is a cloud storage firm that houses its servers in Toronto. The company’s Personal Pro Plan offers five hundred gigabytes of storage for an affordable fifty dollars per year. Occasionally the professor has a need to share his data with his research team. Sync.com would allow him to do this via shareable URLs to the professor’s platform. Sync.com also includes version controlling, something that will be important to the professor as his team starts to sift through and work with the data. Security is also paramount to Dr. Green as he is the only one who can view sensitive interview data. Sync.com would allow the professor to implement access control measures. This would ensure that his research team has access to the data but not to the files containing sensitive information.

Another feature the professor is looking for from his storage solution is accessibility. Presently he keeps a USB drive with him at all times as a fail-safe option. The use of cloud storage would ensure that the professor has access to his data anywhere he goes that has a secure Wi-Fi connection. Sync.com also allows users to access their files via mobile devices as well via a mobile application. If the professor uses a smartphone he can have access to his files at any time regardless of a Wi-Fi connection. This would remove the need for carrying around a USB drive.

**Access, Sharing and Reuse**

The primary user, contributor and manipulator of the data will be Dr. Green. At times the professor will need his research team to access the project’s files. However there are certain audio files which will be restricted for the professor’s exclusive access. Besides a couple of graduate students Dr. Green does not wish to share his data with anyone else at this given time. This means the project’s files will have be protected with the appropriate Creative Commons licensing agreement.

In order to best protect his intellectual property the professor should pursue a CC-BY-NC-ND license. This type of agreement dictates that users of Dr. Green’s data cannot use, modify or commercialize his data without his permission. Specifically this type of Creative Commons license is focused on maximizing the integrity and privacy of the creator’s content. Dr. Green will need this type of license since a primary concern is the proprietary of his data.

In terms of sharing using Sync.com will allow the professor to act as a gatekeeper for his data. Specifically he will have sole authority to share his project data with whoever he chooses. Two-factor authentication and device lockout functionality provided by the platform will further increase his sense of security of access.

It is important to note that while Dr. Green wishes for his data to be proprietary he may run into a dilemma with the university. The university employs and funds the professor so it will be interesting to note who actually owns Dr. Green’s data. There is a very real possibility that the university may lay claim to the professor’s project, depending on how his employment contract was constructed.

**Archiving**

Since the professor’s project will span a number of years, he will need an archive system as he looks to refer his previous work. Building off that Dr. Green should refer to a few best practices of archiving. First he should ensure that he is regularly backing up his files that he has stored on his laptop/desktop. As new data comes in it can become an afterthought to store it on the cloud rather than just your computer. Essentially Dr. Green will have to get into the habit of saving files on the single cloud server rather than his previous practices.

Another recommendation for Dr. Green is to have one of his members of his research team digitize his his spreadsheets from his work prior to 2002. It would be worth archiving this information as the professor can easily refer back to this work. He could store the digital files via Sync.com, as his cloud drive will have ample space to support them.

The final aspect of archiving concerns duration of the data’s usefulness or specifically its lifecycle. There are six stages in the data lifecycle; creating, processing, analyzing, preserving, access to and re-using data. The stages most predominant with Dr. Greens work are preserving, access and re-using data. Preservation of data pertains to the proper backup and storage of Dr. Green’s files. Following best practices of constant cloud-based backups would ensure this step is adhered to. Secondly access to data and the concerns surround who will have access is important to the professors work. Due to confidentiality the professor needs restrict access to his files. Lastly re-using data will be important for the purposes of study recreation. If the professor ever decides to release his study he will have to make sure the data associated with it is within arms reach for the purposes of reusing. This will be important as colleagues look to validate Dr. Green’s study by recreating his project.

**Conclusion**

Dr. Green’s research will need a unique solution as he has many different streams of data. Open-source healthcare stats, interviews and quantitative excel sheets all have different information pertaining to his research. The professor should look to consolidate files by categorizing them within container files such as ZIP files. This would reduce the volume of files, organize them by content and reduce his overall storage needs.

Once his files are consolidated he should then utilize a cloud-based storage platform for housing them. It is important to note that since his data is Canadian he will have to store it in a Canadian platform. Sync.com is a cloud-based Canadian storage company who offers a competitively priced personal plan for fifty dollars annually. This plans includes five hundred gigabytes of data.

Most importantly however this platform offers version and access control. Both of these are paramount to the professor given the level of confidentiality within the data. As the professor’s research begins to grow there will be a greater need to adhere to archiving best practices. The professor is thus encouraged to follow through with the recommendations given in this report throughout the duration of the project.